引入 Altmetrics 指标的掠夺性期刊识别研究*

汪林梓 1) 章博昕 1) 陈铭 1)**

1)南京大学信息管理学院，江苏省南京市栖霞区仙林大道 163 号 210046

摘 要：[**目的**]探讨将 Altmetrics 指标应用于掠夺性期刊识别的效果，为掠夺性期刊的识别提供新视角。[**方法**]基于 Logit 回归模型，分别构造了仅含平均被引量、仅含 Altmetrics 存在率指标以及同时融合这两个指标的三个掠夺性期刊判别模型，并通过 ROC 曲线对三个模型的拟合效果进行了比较。最后使用邀请投稿邮件中的期刊数据对模型效果进行了验证。[**结果**]同时融合平均被引量和 Altmetrics 存在率的掠夺性期刊判别模型效果最优，两项指标与期刊掠夺性均呈显著负相关。使用邮件邀请投稿的 14 本期刊数据验证发现，超过 85%被识别为掠夺性期刊，说明了模型的合理性。[**结论**]Altmetrics 指标应用在掠夺性期刊的识别中具有良好的效果，为掠夺性期刊识别提供了新的有益补充。

关键词：掠夺性期刊；Altmetrics；被引频次；Logit 模型；识别研究

0 引言

开放获取(Open Access，OA)是一种为促进学术信息资源共享，推动学术交流而提出的学术期刊出版模式[1]。OA 期刊作为开放获取的主要产物，向公众读者免费开放已发表的文章，促进了学术成果更为高效、便捷的传播。OA 期刊出版商需要向作者收取相应的文章出版费用以保持运营与盈利，然而在此过程中有些不良出版商利用了这一模式，大批量的、不加审核的出版付费文献，以实现利润收入最大化。美国科罗拉多大学的图书馆员 Beall 将这样利用开放获取模式进行伪造与虚假承诺，不负责任地收取作者的文章出版费用以获利的期刊，称为掠夺性期刊[2]，掠夺性期刊随即出现在学者们的视野中并受到广泛关注与讨论。

掠夺性期刊利用开放获取的特征，掩盖其欺骗伪造的本质，谋取私利，对开放获取的运动产生一定干扰与破坏[3]。近年来，掠夺性期刊呈现快速增长趋势，严重危害学术生态，Nature 在 2022 年 3 月发表的数据表明，掠夺性期刊的数量已超过 15500 种并且快速增长[4]。然而，目前学术界对掠夺性期刊的识别研究还比较薄弱。现有识别掠夺性期刊的方法主要依赖列表式方法，如 Beall 创建的“potential, possible, or probable predatory scholarly open-access publishers”清单及 Kscien 掠夺性名单委员会（PLC）建立的 Kscien’s list[5]。但这类列表法存在明显局限：一是生成和更新困难，新出现的掠夺性期刊难以被及时收录；二是难以覆盖所有掠夺性期刊, 存在一定漏判率。因此随着掠夺性期刊文章的出版数量逐年增多，如何识别、判断掠夺性期刊成为迫切需要关注的问题。近年来，Altmetrics 作为衡量学术影响力的新型指标受到广泛关注，也为期刊评价提供了新的视角。本研究拟利用 Altmetrics 指标判断期刊掠夺性，以期建立更为有效的识别方法。

1 文献综述

1.1 掠夺性期刊特征与识别方法研究

鉴于掠夺性期刊在全球范围带来的巨大影响与危害，学者们从不同维度研究了掠夺性期刊不同与其他期刊的特征。2022 年国际科学院组织（IAP）发布的《打击掠夺性期刊和会议》（Combatting predatory academic journals and conferences）报告[6]中利用了图谱方法，对各类期刊进行了特征的具体区分，其中掠夺性期刊的典型特征包括不存在或不正确的同行评审、模仿其他期刊或网站、无编委或假编委、替代或虚假影响因子等。Shamseer 等[7]曾指出，英文掠夺性科技期刊具有 13 种特征，其中包括网站拼写和语法错误、承诺快速发表、没有撤回政策等。Frandsen[8]发现在掠夺性期刊中发文越多的作者，他们在整个 Scopus 中的平均发文量（以及中位数）都更高，呈现出一定的正相关关系。根据多种

*本文系江苏高校哲学社会科学研究重大项目“开放环境下学术期刊的信誉风险预警研究”（项目编号：2022SJZD093）研究成果。

**作者简介：汪林梓（ORCID：0009-0009-9092-5915），硕士研究生，linzi.wang@smail.nju.edu.cn；章博昕（ORCID：0009-0006-6749-3358），硕士研究生；

通讯作者：陈铭（ORCID：0000-0001-5061-6821），博士，副教授，硕士生导师，chenming@nju.edu.cn

维度的研究，发现掠夺性期刊在费用、宣传、审理及政策等多方面都存在一定的漏洞与问题，这也成为了掠夺性期刊较为明显的特征。

随着掠夺性期刊特征不断地被发现与总结，近年来研究人员对掠夺性期刊的识别指标与方法也相应地提出了许多观点。针对掠夺性期刊的出版高昂费用，Xia[9]通过研究掠夺性期刊的收费情况，发现文章处理费（APC）不能作为区分期刊是否为掠夺性期刊的唯一标准。而王凌峰等[10]通过提出评价学术期刊出版费用合理水平的期刊 JPI 指数，从版面费角度为精确界定掠夺性期刊提供了一个简便有效的客观方法。Ruiter-Lopez L 等[11]从期刊编委会角度出发，使用定量的方法检查了掠夺性期刊的编委会情况，发现其中大部分为高水平学者，发文量中位数为 43，引用次数为 664，h 指数为 14，这说明通过检查编辑部成员情况难以识别期刊是否具有掠夺性。对于期刊网站与用词方面，Chen L-X 等[12]使用机器学习方法从主流掠夺性期刊网站和普通期刊网站中提取出来网站文本内容、关键词等特征，提出了一种基于新模型的掠夺性期刊分类系统。而 Chen 等[13]又发现通过差异评分衡量期刊之间特定词频的差异，提高单词袋模型和 TF-IDF 算法的分类效率，可以帮助识别掠夺性期刊特征词。不过期刊网站和词汇使用上也可以进行调整与修饰，掠夺性期刊与低质量期刊之间的界限仍然比较模糊。Yeo-The 等[14]认为区分掠夺性期刊和真实学术期刊的最重要标准是同行评审的严格性，而且提交作者的动机或意图也至关重要。孔晔晗等[15]从学术出版合法性、商业欺诈行为、学术不端行为等 6 个维度进行对比分析，设置了 18 个二级指标，对识别到的“预警期刊”进行等级划分。从这些动机、相关行为因素上进行分析得出了更深层次的结论，但考虑到一些主观因素较难搜集与判断，这也可能会影响识别结果的准确性。因此，许多学者还会从引文情况对掠夺性期刊进行研究分析。Frandsen[16]追踪了 2013-2016 年 Scopus 中 124 种掠夺性期刊的引用情况，发现这些期刊被引用了 1295 次，每本期刊约 10.5 次，并认为非掠夺性文献中对掠夺性期刊的引用有限。Bo-Christer Björk 等[17]从 Google Scholar 中随机选择了 250 篇在掠夺性期刊上发表的文章，研究了这些文章五年内的被引用数据，发现每篇文章平均被引用 2.6 次，其中 56% 的文章根本没有被引用。

1.2 Altmetrics 指标价值及相关应用

虽然传统引文指标经常被用于衡量学术成果的影响力，但其存在着时间滞后、负面引用及自引等问题，并不能全面的反映出学术成果的影响力[18]。随着社交在线媒体平台的不断发展，科研交流日益网络化，学术活动逐渐开放化，越来越多的科研人员开始使用在线学术网络平台以及社交媒体平台等获取、传播学术资源，并开展学术交流。2010 年 J.Priem 于 Twitter 上首先提出 Altmetrics 这一概念[19]，用来评估学术论文在社交网络上的影响力，通过对学术论文等学术成果在网络中传播交流的追踪，来测度其社会影响力。

Altmetrics 作为新兴的研究影响力指标，自提出便受到广大学者的关注与研究。现有文献显示，关于 Altmetrics 的应用研究多集中在论文或期刊影响力综合评价中。例如，在论文方面，王艳波等[20]提出将用以衡量社会影响力的 Altmetrics 指标和用于衡量学术影响力的传统引文指标相结合，构造出更综合、全面的评价学术论文影响力的指标体系。在期刊维度，王凯利等[21]融合引文分析和 Altmetrics 方法，构建了期刊影响力评价体系，并运用于国际图书情报领域期刊的影响力分析；俞征鹿等[22]人基于 Altmetrics 提及次数指标，对中国英文科技期刊社会影响力进行统计分析。上述研究表明，Altmetrics 指标对论文或者期刊的影响力综合评估的有效性已得到初步验证。

1.3 使用 Altmetrics 指标进行掠夺性期刊识别

尽管使用 Altmetrics 指标来识别掠夺性期刊的研究较少，但初步研究表明其具有应用潜力，笔者在之前的研究中曾分析过图书情报领域中掠夺性期刊与非掠夺性期刊在 Altmetrics 上的表现，发现在图书情报领域中，掠夺性期刊的 Altmetrics 存在率比非掠夺性期刊的低很多[23]。但因掠夺性期刊数据量的限制，未能明确得出 Altmetrics 可以有效识别掠夺性期刊指标的结论，其效用还有待大样本系统研究验证。基于上述情况，本文在已有研究的基础上，选取掠夺性期刊分布广泛的生物医学领域开展研究，在传统基于引文指标识别掠夺性期刊的基础上引入了 Altmetrics 指标，采用 Logit 回归模型构建了掠夺

性期刊的判别模型，为掠夺性期刊的识别提供新的指标和方法。

2 研究设计

2.1 研究假设

基于前文对期刊掠夺性和 Altmetrics 指标及被引指标的文献综述和分析，本研究构建以下两个假设，并拟利用 Logit 回归模型，以期刊的被引指标和 Altmetrics 指标值为自变量，期刊掠夺性作为因变量，验证这两个假设。

假设 1：被引指标与期刊掠夺性具有负相关关系，即被引平均指标越高时，期刊被判定为掠夺性的概率越小。

假设 2：Altmetrics 指标与期刊掠夺性也具有负相关关系，即 Altmetrics 指标越高时，期刊被判定为掠夺性的概率越小。

2.2 研究方法

Logit 模型，也被称为“评定模型”或“分类评定模型”，是离散选择法模型的一种，用于预测事件发生的概率，主要分为二元 Logit 回归模型和多元 Logit 回归模型两类。二元 Logit 回归模型的因变量为二分类，常定义事件发生为“1”，未发生为“0”。多元 Logit 回归模型适用于因变量有多个类别的情况。Logit 模型广泛应用于社会学、生物统计学、计量经济学等领域中。在图情领域中，任海芝等[24]使用多元 Logit 回归模型对图书出版企业官微传播力的影响因素进行了实证分析。本研究主要探讨掠夺性期刊的识别问题，而二元 Logit 模型能够很好的刻画 “掠夺性期刊”及“非掠夺性期刊”这两种类型，由于本文假设被引指标和 Altmetrics 指标存在负相关关系，因此选择使用二元 Logit 模型进行期刊类型的识别分析。

Logit 模型是由传统线性回归模型衍生而来，在多元线性回归模型中 $y=\alpha_0+\sum\alpha_j x_j+\varepsilon$中 y 的取值范围是$(-\infty,+\infty)$，而事件是否发生概率范围为[0,1]。因此，需要对线性回归模型进行 Logit 变换。首先，引入发生比 Odds，Odds 表示事件发生概率和事件不发生概率的比值，Odds 的计算方法如式 1 所示，此时 Odds 的取值范围是$[0,+\infty)$。

$$\text{Odds}=\frac{Probability\ of\ event}{Probability\ of\ no\ event}=\frac{P}{1-P} \quad (1)$$

其中，p 表示事件发生的概率。

将发生比 Odds 取自然对数得到 Logit，这时就可以将 P 从[0,1]映射为$(-\infty,+\infty)$，从而可以进行多元线性回归建模。从概率 P 到 Odds 再到 Logit 的过程，称为 Logit 变化，见式 2：

$$y=\text{logit}(P)=\ln(\text{Odds})=\ln\left(\frac{p}{1-p}\right)=\alpha_0+\sum\alpha_j x_j+\varepsilon \quad (2)$$

其中，p 表示因变量某一取值的概率，x_j为自变量，α_j为自变量的系数，α_0为截距项，ε表示误差项。

2.3 变量定义及数据获取

2.3.1 因变量定义及数据来源

本研究中采用期刊类型二分类离散变量作为因变量，并定义掠夺性期刊=1，非掠夺性期刊=0。Shen 和 Bjork[25]研究中发现，掠夺性期刊的文章数量逐年迅速增加，从 2010 年的 53,000 多篇上升至 2014 年 420,000 多篇，活跃掠夺性期刊约 8000 本，这些掠夺性期刊文章绝大多数来自于生物医学等领域。因此，本研究从生物医学领域随机抽取掠夺性期刊和非掠夺性期刊各 100 本作为研究样本。掠夺性期刊数据来源于 Kscien’s 列表，非掠夺性期刊数据来源于 DOAJ 目录。DOAJ 是由瑞典隆德大学图书馆创建和维护的收录经同行评议的开放存取（OA）期刊目录网站，具有严格的期刊收录标准和期刊评估流程[26]。由于 Kscien’s 列表与 DOAJ 均为开放获取期刊，因此本文抽取的两类期刊样本具有可比性。

2.3.2 自变量定义及数据来源

被引次数在一定程度上反映了文章或者期刊的学术影响力，是评价论文质量和价值的重要指标，也是计算期刊影响因子的核心指标。为消除期刊文章数量差异对被引量的影响，本文使用期刊平均被引数来衡量被引情况，并用变量x_1表示。通过爬取期刊网站收集到2012-2022年期刊所发表的文章数量，并使用Web of Science查询该期间的期刊被引数量。具体计算公式如式3：

$$期刊平均被引数（x_1）=\frac{期刊文章被引总数}{期刊发表文章数} \quad （3）$$

Altmetrics作为衡量学术成果的社会影响力的新指标，拓展了基于引文的传统影响力评估，能更加全面地反映学术成果在数字网络中的影响。自Altmetrics概念提出后，涌现出多种测量工具，如Altmetric.com、 PlumX和Crossref Event Data(CED) 等。其中Altmetric.com信息源广泛，已涵盖全球5000多家主流媒体和1.5万个学术及非学术博客[27]，且向研究人员提供免费使用的机会，因此，近年来大多数Altmetrics的相关研究采用了Altmetric.com工具[28]。本研究通过Altmetric. com的Altmetric Explorer工具，使用期刊ISSN作为检索条件，获取2012-2022年期刊的Altmetrics数据。为消除期刊规模差异影响，本文定义Altmetrics存在率指标，并用自变量x_2表示，具体计算公式如式4：

$$\text{Altmetrics}存在率（x_2）=\frac{存在\text{Altmetrics}得分的文章数}{期刊发表文章数} \quad （4）$$

3 实证分析

3.1 描述性统计

从收集的200本掠夺性期刊及非掠夺性期刊的Altmetrics存在率和平均被引数的描述性统计量（表1）可以观察到，这些掠夺性期刊的Altmetrics存在率平均值仅为0.0088，在这100本期刊中，有81本的Altmetrics存在率都为0，即81%掠夺性期刊的文章没有Altmetrics分数，这表明了掠夺性期刊文章较少受到社交媒体的关注与讨论。这些掠夺性期刊的平均被引数均值为0.7210，即平均每篇文章被引用约0.7210次。

相比掠夺性期刊，非掠夺性期刊的Altmetrics存在率平均为0.338，远大于掠夺性期刊的0.0088。这表明了非掠夺性期刊在社交网络上的影响力远大于掠夺性期刊。对于非掠夺性期刊，仅有23本期刊的Altmetrics存在率为0，并且Altmetrics存在率最高的期刊能够达到0.9916，表明与掠夺性期刊相比，非掠夺性期刊在社交网络上受关注程度和影响力更大。在期刊被引方面，非掠夺性期刊平均被引数的均值为5.1763，高于掠夺性期刊。侧面反应了掠夺性期刊文章的质量难以保证，未必能提供建设性观点，较少被引用，而非掠夺性期刊的文章质量可能更高，对科研人员的研究更有实质的帮助，所以被引用频次更多。

表1 掠夺性期刊及非掠夺性期刊数据描述性统计分析

变量	期刊类型	最小值	最大值	平均数	标准差
Altmetrics存在率	掠夺性期刊	0.000	0.2312	0.0088	0.0337
	非掠夺性期刊	0.000	0.9916	0.3357	0.3264
平均被引数	掠夺性期刊	0.000	12.6412	0.7210	1.7353
	非掠夺性期刊	0.000	22.0481	5.1763	5.6598

3.2 基于Logit模型的实证分析结果

由上述分析可得，平均被引数和Altmetrics存在率在掠夺性期刊和非掠夺性期刊上存在着较大差异，那我们是否可以把它们作为判定期刊掠夺性的指标呢？笔者使用Stata14.0软件，建立和比较了以下三个Logit回归模型。首先基于引文指标构造掠夺性期刊识别模型（模型1），然后探讨使用Altmetrics存在率构建的判别模型（模型2）效果，最后结合平均被引数和Altmetrics存在率构建掠夺性期刊识别模型（模型3），并对这三个模型及效果进行了比较分析。具体模型及效果分析如下。

3.2.1 仅针对引文指标的判别模型（模型1）分析结果

通过使用Stata14.0软件对期刊的平均被引数进行Logit模型回归分析，得出结果如表

2 所示。Logit 模型拟合结果的表达式为 $y=\ln\left(\frac{p}{1-p}\right)=0.864-0.438x_1$。可以看出，平均被引数变量系数为-0.438<0（sig<0.05），这表明了平均被引数对期刊是否为掠夺性期刊呈显著负相关，说明假设 1 成立，即当平均被引数较低时，期刊为掠夺性期刊的概率更大。

表 2 针对引文指标 Logit 模型回归结果

变量名称	系数	Z 值	Sig
常数项	0.864	4.46	0.000
平均被引数	-0.438	-5.13	0.000

在对模型参数计算完成后，需要对模型计算出的预期概率和实际概率是否能够有效的拟合做出评价。如果实际观测值能够与模型预测值有着较高的一致性，则认为该模型能够拟合数据；反之，则不能接受该模型，需要重新设置模型的变量。在本研究中笔者采用 Hosmer-Lemeshow 检验对二元 Logit 回归模型的拟合优度进行检验。

Hosmer-Lemeshow 检验表示拟合值和观测值的吻合程度。如 Sig<0.05，表明模型的预测值与观测值存在显著差异，评分模型工作效果欠佳；反之，Sig>0.05 则认为在可接受的水平上模型的估计拟合了数据，表明评分模型工作效果良好。对引文指标的判别模型进行 Hosmer-Lemeshow 检验，得出检验的 Sig 值为 0.000<0.05，这一结果说明了仅依据平均被引量建立的 Logit 判别模型的拟合结果欠缺，为此我们尝试引入 Altmetrics 指标进行判别分析。

3.2.2 仅针对 Altmetrics 的判别模型（模型 2）分析结果

首先，仅对期刊的 Altmetrics 存在率进行 Logit 判别模型建立，得到 Logit 模型拟合表达式为 $y=\ln\left(\frac{p}{1-p}\right)=1.036-14.871x_2$，建模结果如表 3 所示。可以看出 Altmetrics 存在率的系数为-14.871<0（sig<0.05），表明了期刊的 Altmetrics 存在率对期刊是否是掠夺性期刊呈显著负相关，说明假设 2 成立，即期刊的 Altmetrics 存在率越高，期刊为掠夺性期刊的概率越小。Altmetrics 存在率能够反映出一本期刊学术成果的社会影响力，因此这一结果也是合理的。该部分继续使用 Hosmer-Lemeshow 检验来对模型拟合程度进行检验，得出使用 Altmetrics 存在率建立的判别模型 Hosmer-Lemeshow 检验的 Sig 值为 0.997>0.05，说明了该模型的拟合效果较好。

表 3 针对引文指标 Logit 模型回归结果

变量名称	系数	Z 值	Sig
常数项	1.036	5.24	0.000
Altmetrics 存在率	-14.871	-4.22	0.000

3.2.3 引入平均被引数和 Altmetrics 存在率的判别模型（模型 3）分析结果

除了单独使用 Altmetrics 存在率构造了 Logit 判别模型外，笔者还建立了同时引入平均被引量和 Altmetrics 存在率变量的 Logit 模型回归模型，Logit 模型拟合结果的表达式为 $y=\ln\left(\frac{p}{1-p}\right)=1.220-0.204x_1-12.015x_2$，回归结果如表 4 所示。其中，被引平均数和 Altmetrics 存在率指标的显著性水平分别为 0.033 和 0.002，均小于 0.05。这表明拟合效果是显著的，平均被引数和 Altmetrics 存在率对期刊是否为掠夺性期刊具有负向影响，假设 1 和假设 2 成立。即当平均被引数和 Altmetrics 存在率较低时，期刊为掠夺性期刊的概率更大。这与模型 1 和模型 2 得到的结果也是相一致的。

表 4 针对引入 Altmetrics 模型回归结果

变量名称	系数	Z 值	Sig
常数项	1.220	5.51	0.000
平均被引数	-0.204	-2.13	0.033
Altmetrics 存在率	-12.015	-3.15	0.002

利用 Hosmer-Lemeshow 检验对模型拟合优度进行检验，得到结果为 Sig=0.3568>0.05。

Hosmer-Lemeshow 检验中，当 Sig 值大于 0.05 的时候拟合效果较好，当 Sig 值大于 0.1 的时候拟合效果更佳，说明引入 Altmetrics 存在率后的判别模型拟合效果较好。

3.2.4 三个模型比较分析结果

ROC 曲线下的面积来检验模型预测准确率的情况，并认为当曲线下面积大于 0.75 时，该模型有足够的辨别力。根据曲线位置，ROC 曲线把整个图划分成了两部分，曲线下方部分的面积被称为 AUC（Area Under Curve），用来表示预测准确性，AUC 值越高，也就是曲线下方面积越大，说明预测准确率越高。曲线越接近左上角（X 越小，Y 越大），预测准确率越高。本文使用 ROC 分析来进行三个模型的比较分析。

表 5 模型 ROC 比较

模型	AUC 面积	标准误	渐进 95%的置信区间	
			上限	下限
模型 1	0.8204	0.0292	0.76314	0.87766
模型 2	0.8468	0.0260	0.3780	0.7559
模型 3	0.8908	0.0226	0.84649	0.93511

从模型 ROC 结果（表 5）可以看出，模型 3 的预测效果要优于模型 1 和模型 2，即模型 3 的预测准确率更高。这也就是说，在判别模型的构建上，同时融合了平均引文数和 Altmetrics 存在率的模型要比单独利用平均引文数、Altmetrics 存在率的模型效果要好，单独使用 Altmetrics 存在率构造的判别模型又比单独使用平均引文数的模型效果更好，进而证明了本文提出引入 Altmetrics 指标进行掠夺性期刊判别的合理性和正确性。

4 模型验证

为了对模型效果进行验证，笔者收集了在近一年内所收到的邀请投稿邮件中的期刊数据。Sureda－Negre 等[29]在对西班牙一所大学教育领域的三位教授在 3 个月内收到邀请在期刊上发表文章的电子邮件中的期刊进行分析发现，大多数的期刊（69.7%）在掠夺性期刊名单中，并认为通过邮箱向学者发送投稿邀请的期刊大部分质量不高。因此，我们认为选择邮箱中的邀请发文期刊进行判别模型效果验证是合理的，在剔除了在 Kscien’s list 中出现的期刊后，得到期刊如表 6 所示。

表 6 邀请投稿邮件中的期刊信息

期刊名称	期刊网址	ISSN Online
Review of Contemporary Business Research	http://rcbrnet.com/	2333-6420
American Journ a l of Information Science and Technology (AJIST)	http://www.amjist.org/cggh6j/zi8vz	2640-0588
Journal of Intercultural Communication (JICC)	https://immi.se/intercultural	1404-1634
Journal of Management Information System and E-commerce	http://jmisenet.com/	2373-7638
International Journal of Business & Economic Development	https://ijbed.org/	2692-5583
International Journal of Business and Applied Social Science	www.ijbassnet.com	2469-6501
International Journal of Library and Information Studies	https://www.ijlis.org/	2231-4911
Journal of Economics & Management Research	http://dspace.lu.lv/dspace/handle/7/2553	2755-0214
International Journ a l of Sustainability Management and Information Technologies	http://www.ijosmit.com/se4bqw/j7IeA	2575-5110
Health Informatics Journal	https://journals.sagepub.com/home/JHI	1741-2811
International Journal of Business and Social Science Research	www.ijbssrnet.com	2709-2143

International Journal of Sustainability Management and Information Technologies	https://www.sciencepublishinggroup.com/j/ijsmit	2575-5110
Journal of Business and Social Science Review (JBSSR)	jbssrnet.com	2690-0874
International Journal of Business & Management Studies (IJBMS)	https://ijbms.net/	2694-1449

使用相同的数据收集方法对这些期刊的引文数据和 Altmetrics 数据进行收集，并将其带入模型 3 中进行验证，最终得到结果如表 7 所示，可以看到在这 14 本验证期刊中仅有 1 本期刊 *Health Informatics Journal* 为掠夺性期刊的预测概率为 0.003<0.5，笔者查阅后发现，期刊 *Health Informatics Journal* 为 SCI 期刊，JCR 分区为 Q3 区，这同我们所预测的该期刊为掠夺性期刊的概率是相匹配的。除此之外余下 13 本的预测概率大于 0.5，12 本预测概率大于 0.7，这说明了这些期刊虽然还没有出现在 Kscien 掠夺性期刊的列表中，但它们为掠夺性期刊的概率较高，这也和 Sureda－Negre 等[29]提出的通过邮箱向学者发送投稿邀请的期刊质量不高的观点相符合。也表明了融合了平均引文数和 Altmetrics 存在率的掠夺性期刊判别模型是合理的。

表 7 模型验证结果

期刊名称	Altmetrics 存在率	平均被引数	预测概率
Review of Contemporary Business Research	0.0000	0.5870	0.7504
American Journ a l of Information Science and Technology (AJIST)	0.0000	0.0000	0.7721
Journal of Intercultural Communication (JICC)	0.0714	1.3524	0.5215
Journal of Management Information System and E-commerce	0.0000	0.4583	0.7552
International Journal of Business & Economic Development	0.0000	0.1289	0.7675
International Journal of Business and Applied Social Science	0.0000	0.0841	0.7691
International Journal of Library and Information Studies	0.0000	0.1117	0.7681
Journal of Economics & Management Research	0.0000	0.0337	0.7709
International Journ a l of Sustainability Management and Information Technologies	0.0000	0.0000	0.7721
Health Informatics Journal	0.6297	9.1725	0.0003
International Journal of Business and Social Science Research	0.0000	0.1304	0.7674
International Journal of Sustainability Management and Information Technologies	0.0000	0.5111	0.7533
Journal of Business and Social Science Review (JBSSR)	0.0000	0.0168	0.7715
International Journal of Business & Management Studies (IJBMS)	0.0000	0.1282	0.7675

5 讨论和结论

本文基于 Logit 回归模型，分别构建了仅含期刊的平均被引量，仅含 Altmetrics 存在率指标，以及同时包含期刊的平均被引量和 Altmetrics 存在率指标的掠夺性期刊判别模型。通过模型比较分析后发现，仅含 Altmetrics 存在率指标与同时引入平均被引量和 Altmetrics 存在率的掠夺性期刊判别模型效果较优。虽然传统认为期刊的引文指标对于期刊学术影响力的衡量较为适用，但 Oviedo-García 对掠夺性期刊出版商 MDPI 的分析显示，某些掠夺性期刊存在较高的自引率[30]，导致部分掠夺性期刊有着较高的被引量。这说明，这些期刊的高被引量存在被操纵的可能，纯粹以引文指标来判断期刊是否为掠夺性期刊并不

准确。在 Web2.0 环境下，Altmetrics 指标充分利用了学术社交网络进行文献计量，Altmetrics 指标因为数据更新更加及时，进而能够避免学术成果影响力评估的滞后性，并可以补充性地反映学术成果的社会影响力。Altmetrics 指标和引文指标共同应用，可以更全面地评估学术成果的影响力。在本研究中，同时引入平均被引量和 Altmetrics 指标构建的判别模型，综合了两类指标的优势，因此对判断期刊的掠夺性具有更好的效果。

在此基础上，笔者收集了邮箱中投稿邀请邮件的期刊数据，将其带入同时引入平均被引量和 Altmetrics 存在率的掠夺性期刊判别模型，对该模型进行验证。结果表明，邮箱邀请投稿的期刊很大概率上为掠夺性期刊。这表明了 Altmetrics 指标在识别掠夺性期刊方面具有良好的效果，可以较好地判断期刊的良劣，为掠夺性期刊的识别提供新的指标和方法。

不可否认的是，本研究仍然存在一些局限性。首先，由于数据获取困难，我们仅能初步验证使用期刊被引指标和 Altmetrics 指标识别掠夺性期刊的有效性；其次，文章中应用的方法还有待优化。在未来的研究中，我们将尝试从以下两个方面进行改进：一是扩大指标类型，加入更多能够有效识别掠夺性期刊的指标，以丰富模型输出；二是增加样本量，并尝试采用更为前沿的方法，如机器学习中的随机森林模型，以提升识别效果和模型精度。总体而言，本研究对于掠夺性期刊的识别具有一定启发意义，但仍需在数据和方法上做进一步拓展。

参考文献

[1] 李武，刘兹恒. 一种全新的学术出版模式:开放存取出版模式探析［J］. 中国图书馆学报，2004（6）： 67-70.

[2] Beall J. Predatory journals and the breakdown of research cultures[J]. Information development, 2015, 31(5): 473-476.

[3] 戴琦,袁曦临.开放获取期刊的学术声誉风险及其预警研究[J].中国科技期刊研究,2018,29(11):1063-1071.

[4] Nature ： 全 球 掠 夺 性 期 刊 已 超 15500 种 [EB/OL].[2022-08-16]. https://mp.weixin.qq.com/s/WJ_4URZ2MqUVhjuzl2Jm-g.

[5] Kakamad F H, Mohammed S H, Najar K A, et al. Kscien's list; a new strategy to hoist predatory journals and publishers[J]. International Journal of Surgery Open, 2019, 17: 5-7.

[6] The InterAcademy Partnership. Combatting predatory academic journals and conferences[EB/OL].[2023-08-21].https://www.interacademies.org/project/predatorypublishing.

[7] Shamseer L,Moher D,Maduekwe O,et al. Potential predatory and legitimate biomedical journals: Can you tell the difference? A cross-sectional comparison[J]. BMC Medicine,2017,15(1):1-14.

[8] Frandsen, TF. Authors publishing repeatedly in predatory journals: An analysis of Scopus articles[J]. Learned Publishing,2022,35(4): 598-604.

Xia J. Predatory journals and their article publishing charges[J]. Learned Publishing, 2015, 28(1): 69-74.[10]王凌峰,杜艳玲,唐碧群.国内“掠夺性期刊”量化定义:期刊 JPI 指数[J].晋图学刊,2018(03):5-9.

[9] Ruiter-Lopez L, Lopez-Leon S, Forero D A. Predatory journals: Do not judge journals by their Editorial Board Members[J]. Medical Teacher, 2019, 41(6): 691-696.

[10] Chen LX, Wong KS, Liao CH, et al. Predatory Journal Classification Using Machine Learning[C]. 2020 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII). IEEE, 2020: 193-196.

[11] Chen, LX., Su, SW., Liao, CH. et al. An open automation system for predatory journal detection[J]. Scientific Report,2023(13) : 2976.
[12] Yeo-Teh NSL, Tang BL. Wilfully submitting to and publishing in predatory journals - a covert form of research misconduct?[J]. Biochemia medica, 2021(3) : 0-0.
[13] 孔 晔 晗 , 彭 琳 , 袁 军 鹏 , 于 艺 浩 . 科 技 期 刊 预 警 监 测 系 统 构 建 [J]. 编 辑 学 报,2022,34(04):369-377.
[14] Frandsen T F. Are predatory journals undermining the credibility of science? A bibliometric analysis of citers[J]. Scientometrics, 2017, 113(3): 1513-1528.
[15] Björk B C, Kanto-Karvonen S, Harviainen J T. How frequently are articles in predatory open access journals cited[J]. Publications, 2020, 8(2): 17.
[16] 邱均平,余厚强.替代计量学的提出过程与研究进展[J].图书情报工作,2013,57(19):5-12.
[17] Priem J. I like the term #articlelevelmetrics, but it fails to imply * diversity* of measures . Lately, I'm liking #altmetrics [EB/OL]. ［ 2023-08-20 ］ . https: //twitter.com/jasonpriem/status/25844968813.
[18] 王艳波,汪雷,吴小兰.基于 Altmetrics 与传统文献计量指标的国内学术论文影响力评价研究[J].情报理论与实践,2020,43(06):76-81.
[19] 王凯利,黄晓,吴江.融合 Altmetrics 和引文分析的期刊影响力综合评价研究[J].情报科学,2021,39(02):169-177.
[20] 俞征鹿,马峥,田瑞强等.中国英文科技期刊国际社会影响力表现——基于 Altmetrics 提及次数指标[J].编辑学报,2021,33(03):349-354.
[21] Chen M, Wang L. An Altmetrics and citation analysis of selected predatory journals in library and information science field[J]. The Journal of Academic Librarianship, 2022, 48(6): 102618.
[22] 任海芝,潘英杰.基于 Logit 模型的图书出版企业官微传播力影响因素研究[J].情报探索,2022(06):62-70.
[23] Shen C, Björk B C. ‘Predatory’open access: a longitudinal study of article volumes and market characteristics[J]. BMC medicine, 2015, 13(1): 1-15.
[24] 王琳辉,倪明.中国开放存取期刊申请加入 DOAJ 数据库策略探析[J].出版与印刷,2022(05):74-83
[25] Altmetric. Attention sources tracked by Altmetric ［EB/OL］. ［2022 － 11 － 24］. https://help.altmetric.com/support/solutions/articles/6000235983-attention-sources-tracked-by-altmetric
[26] Ortega J L. Altmetrics data providers: A metaanalysis review of the coverage of metrics and publication[J]. El profesional de la información (EPI), 2020, 29(1):3.
[27] Sureda‐Negre J, Calvo‐Sastre A, Comas‐Forgas R. Predatory journals and publishers: Characteristics and impact of academic spam to researchers in educational sciences[J]. Learned Publishing, 2022, 35(4):441-447
[28] Oviedo-García M A. Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI)[J]. Research Evaluation, 2021, 30(3): 405-419.

A study of predatory journal identification using Altmetrics metrics

Abstract:

[Purposes] This article explores the effect of introducing the Altmetrics index into the identification of predatory journals, in order to provide a new perspectives for the identification of predatory journals.
[Methods] Based on the logit regression model, three models for the identification of predatory journals were constructed using the average number of citations alone, the presence rate of altmetrics alone, and a combination of both indicators. The ROC curve was used to compare and analyze the fitting effects of the three models. Finally, journal data in invitation emails were used to validate the model effects.
[Findings]The model integrating both average citations and Altmetrics existence rate performed the best for identifying predatory journals. with both indicators showing significant negative correlation with journal predatoriness. Validation using the data from 14 journals that sent out solicitation emails showed that over 85% of them were identified as predatory journals by the model, demonstrating the rationality of the model.
[Conclusions] The application of the Altmetrics metric has shown good results and provides valuable new perspectives in the identification of the predatory journals.
Key words:　Predatory journals; Altmetrics; Citation analysis; Identification studies

[作者贡献说明]：汪林梓：研究思路设计，数据收集整理，论文撰写及论文修改；
章博昕：数据收集整理，论文撰写及论文修改；
陈铭：研究思路设计，论文撰写及论文修改。